

Discussion Papers

916

Nicolas R. Ziebarth

**Measurement of Health, the Sensitivity
of the Concentration Index,
and Reporting Heterogeneity**

Berlin, August 2009

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2009

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Available for free downloading from the DIW Berlin website.

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN.
Papers can be downloaded free of charge from the following websites:

http://www.diw.de/english/products/publications/discussion_papers/27539.html
<http://ideas.repec.org/s/diw/diwwpp.html>
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=1079991

Measurement of Health, the Sensitivity of the Concentration Index, and Reporting Heterogeneity[‡]

Nicolas R. Ziebarth
SOEP at DIW Berlin and TU Berlin*

June 18, 2009

[‡]I take responsibility for all errors in and shortcomings of the article and would like to thank participants at various seminars for their helpful comments and discussions. Financial support in the form of a scholarship from the *Stiftung der Deutschen Wirtschaft* (sdw, Foundation of German Business) is gratefully acknowledged.

*German Institute of Economic Research (DIW Berlin), Socio-Economic Panel Study (SOEP), Mohrenstraße 58, 10117 Berlin, Germany, and Berlin University of Technology (TU Berlin), phone: +49-(0)30-89789-587, fax: +49-(0)30-89789-109, e-mail: nziebarth@diw.de

Abstract

Using representative survey data from the German Socio-Economic Panel Study (SOEP) for 2006, we show that the magnitude of such health inequality measures as the concentration index (CI) depends crucially on the underlying health measure. The highest degree of inequality is found when dichotomized subjective health measures like health satisfaction or self-assessed health (SAH) are employed. Measures of medical care usage like doctor visits result in substantially lower concentration indices. Moreover, with the use of SF12, a generic health measure, the inequality indicator is reduced by a factor of ten. Scaling SAH by means of the SF12 leads to similar results to those with the pure SF12 measure. Employing generic health measures used with other populations like the Canadian HUI-III or the Finish 15D to cardinalize SAH has a significant impact on the degree of inequality measured. Finally, by contrasting the physical health component of the SF12 to the unambiguously objective grip strength measure, we provide evidence of the presence of income-related reporting heterogeneity in generic health measures.

Keywords: health measures; health inequality; SF12; grip strength; German Socio-Economic Panel Study (SOEP)

JEL classification: D30; D31; D63; I10; I12

1 Introduction and Background

During the last decades, measuring health inequalities has become increasingly popular. Major reasons are certainly an increased data availability, an increased technological capacity to handle the data, and more refined methods to analyze the data.

Although research on health inequalities has increased in recent years, it is still an underdeveloped field as compared to the field of income inequalities. The public focus on income inequalities seems to blur the awareness of the probably more fundamental issue of equitable health distributions. Work on income-related health inequalities links the two research fields.

Within the research field of inequalities in health, one can define several subfields. In order to sharpen the focus and to reveal the contributions of the study at hand, we categorize the studies into five main inter-related fields (see Figure 1). *Field A* includes all studies that deal with health measurement. Health measures can be classified as:

- a) **Subjective health measures** such as health satisfaction or self-assessed health (SAH): SAH is the most popular health measure and most studies on health inequalities in developed countries rely on it. Analogously to what advocates of happiness economics claim, one can take the view that differences in health *satisfaction* should be the relevant indicator rather than differences in true health.
- b) **Medical care measures** like the number of doctor visits, the number of days in hospital, or the number of sick leave days.
- c) **Generic health measures** such as the SF12, the HUI-III, or the 15D: these preference-based measures are presumably objective and comprehensive. They are generated from self-reported questions which cover various health dimensions and are aggregated into a single index. Henceforth, we refer to them as quasi-objective.
- d) **Vignette-based health measures**: respondents are asked to rate their own health and those of fictitious individuals. From this, the respondent's "true" health status is derived.

- e) **Objective health measures** such as height at a certain age or grip strength: they are based on diagnosis by physicians or use invasive or non-invasive procedures.

Using data that incorporate an extraordinarily rich set of different health measures, our work predominately contributes to *Field A*. The main aim of the paper is to analyze the sensitivity of the concentration index with respect to the health measure used. In addition, we show that cardinalizing SAH by means of continuous, generic health distributions has an impact on the size of the inequality measure. We perform this exercise by comparing the German SF12, the Canadian HUI-III, and the Finish 15D as bases of the scaling procedure.

[Insert Figure 1 about here]

Closely related to the question of how to define and measure health states is *Field B*, which deals with reporting heterogeneity. Reporting heterogeneity, which is sometimes also called “state dependent reporting bias,” or “scale of reference bias,” points towards the finding that some population groups may systematically rate their health status as different from others. There is evidence that cultural as well as semantic differences vary in a systematic way with the response behavior (Mathers and Douglas, 1998; Jürges, 2007; Bago d’Uva et al., 2008). While these findings are challenges for cross-country comparisons, the phenomenon of differential reporting also exists *within* countries when samples are stratified by education, age, gender, or income. Older people tend to judge their health more mildly, whereas the findings for gender, education, and income are mixed, depending on the country (van Doorslaer and Gerdtham, 2003; Lindeboom and van Doorslaer, 2004; Etilé and Milcent, 2006; Jürges, 2008; Bago d’Uva et al., 2008). Reporting heterogeneity in particular biases socioeconomic status (SES) related health inequality measures. By definition, reporting heterogeneity might be an issue, and has been proven to exist, when health measures are based on self-reported answers, i.e., in all cases except e) in the above classification.

When talking about reporting heterogeneity and SAH, a useful distinction is the one between “cut-point shifts” and “index shifts.” By index shifts we mean when *all* thresholds of the ordinal SAH measure are shifted parallelly for certain subpopulations. Cut-point

shifting occurs if certain subpopulations value a specific health state differently and hence single SAH cut-off points are shifted. Lindeboom and van Doorslaer (2004) identified the non-separability of true health differences from index shifts as a fundamental identification problem.

Conventional tests on reporting heterogeneity mirror subjective health measures against allegedly objective measures like the HUI-III or vignettes. The main underlying assumption is that preference-based generic health measures and vignettes are not prone to reporting heterogeneity. Since these health measures are derived from health self-reports, this assumption is questionable. One might take the view that solely measures of class e) are appropriate for testing reporting heterogeneity. However, they almost always suffer from a lack of multidimensionality and reduce health to a very narrow physical definition. Hence, researchers face the even more fundamental problem of identifying reporting heterogeneity – should measures of type c) or d) incorporate, at least to some degree, reporting heterogeneity.

Baker et al. (2004) compared self-reports of specific chronic conditions to medical records and found considerable reporting errors which were related to the labor market activity of the respondents. Other studies found similar results (Johnston et al., 2009). While these papers rely on specific diagnosis, we provide evidence of the existence of reporting heterogeneity in generic health measures. This is done by contrasting the physical health component of the SF12 to an unambiguously objective measure of physical health – the grip strength.

So as to not overload the paper, we will only briefly touch on the three other fields, identified as *C*, *D*, and *E* in Figure 1. *Field C* deals with the type of inequality indicator (e.g. Erreygers (2009)), while *Field D* tests the sensitivity of the indicator with respect to the SES measure used (e.g. Wagstaff and Watanabe (2003); Lindelow (2006)). We make use of the concentration index and relate it to the equivalized household income since a broad consensus has been found for these two indicators. Studies in *Field E* deal with the decomposition of inequality measures into its determinants (e.g. Wagstaff and Watanabe (2003); van Doorslaer and Koolman (2004); Lauridsen et al. (2007); Huber

(2008)). However, mechanically applied, the Oaxaca-Blinder type of decomposition merely decomposes the index into its correlation components. We take the view that this exercise is only of tangible value for policymakers if distinct causal relationships between potentially endogenous variables – such as education or income – and health have been proven to exist for the country in question. For Germany, this has only been shown for the effect of income on health satisfaction in the aftermath of reunification (Frijters et al., 2005).

Our work makes various contributions to the literature on health inequalities. To our knowledge there is no study that systematically and rigorously assesses the CI’s sensitivity with regard to ten different health measures and three different SAH scaling procedures. Compared to the other fields in Figure 1 and relative to its immanent importance, there is surprisingly little research on the impact of health measures on health inequality indicators (one recent exception is Tubeuf (2008)). In addition, we show that the choice of the generic health measure to scale SAH has a substantial impact on the size of the inequality measure. Instead of using the Canadian HUI-III, we propose to use the more symmetric German SF12 when European SAH distributions are cardinalized – along with other generic measures – to prove the robustness of the results. Finally, to the best of our knowledge, we are the first to provide evidence of the existence of reporting heterogeneity in generic health measures. In contrast to other studies, our analyses rely on the most recent data which are representative of the most populous European country.

2 Methods

Calculating Concentration Curves and Indices

The concentration curve (CC) is a tool to visualize the degree of socioeconomic inequality in health for the whole distribution. It plots the cumulative share of the population, ranked by a living standard variable, against the corresponding cumulative share of a health measure. The benchmark is the 45°-line which represents the state where health is not systematically related to socioeconomic characteristics. If the health measure on the vertical axis stands for ill-health, then ill-health is more concentrated among the lower

social classes if the CC lies above the line of equality.

Converting the visual representation into a single indicator is achieved by twice defining the area between the CC and the line of equality as Concentration Index (CI). Formally:

$$CI_h = 1 - 2 \int_0^1 CC(r) dr$$

where respondent i 's rank in the income distribution is expressed by r , with $r \in [0, 1]$. The CI lies between -1 and 1 and takes the value zero in the case of no SES-related health inequality. It has negative values if ill-health is more concentrated among the poor. Using micro data, the CI can be computed as

$$CI_h = \frac{2}{n\mu} \sum_{i=1}^n h_i r_i - 1$$

where n is the sample size and μ is the mean of the applied health measure h (Kakwani et al., 1997). Within a regression framework, we obtain the CI by means of

$$2\sigma_r^2 \left(\frac{h_i}{\mu} \right) = \alpha + \beta r_i + \varepsilon_i$$

with σ_r being the variance of the fractional rank.

Since the concentration index is by far the most dominant health inequality indicator and has various advantages over alternative indicators (Wagstaff and van Doorslaer, 2000), the following sections use the concentration index exclusively.

Scaling SAH

It is straightforward to draw CC s and calculate CI s for binary or continuous health measures. However, using an ordinal measure such as SAH is more problematic. Various

methods have been proposed to deal with this issue. The most obvious variant is to dichotomize ordinal measures. On the one hand, dichotomization comes at the cost of loss of information and it has been shown that one might run into problems in cross-country comparisons (Wagstaff, 2005). On the other hand, Etilé and Milcent (2006) have shown that dichotomizing helps to overcome biases associated with reporting heterogeneity.

In a seminal paper, van Doorslaer and Jones (2003) proposed an interval regression based method to cardinalize ordinal health measures like SAH. They used the Canadian HUI-III as an objective and comprehensive measure of health to scale SAH. Subsequent work has confirmed that this method outperforms alternative scaling procedures (Lauridsen et al., 2004). Today, the van Doorslaer and Jones’ method is state-of-the-art but it rests on two strong assumptions: firstly, it is assumed that the underlying generic health measure represents the true health status and is free of systematic reporting errors. Secondly, since most surveys do not include generic health measures – if generic health measures are included, then scaling SAH becomes superfluous – the Canadian HUI-III distribution is often imposed on the population of interest assuming that the Canadian health distribution represents the health distribution of any given country.

After having analyzed the sensitivity of the CI with respect to the underlying health measure, we will look at the sensitivity when different generic health measures are used to scale SAH. Table 1 shows summary statistics of three generic health measures by SOEP SAH categories. Only the SF12 is carried out in the course of the SOEP. The HUI-III values were collected in the course of the Canadian National Population Health Survey (NPHS) 1994/1995 and the 15D measure stems from the Finnish Health Care Survey 1995/1996. All three indices lie between zero and one. The latter two measures are used to simulate the effects of the adoption of a generic health distribution from a foreign population to the population of interest.

[Insert Table 1 about here]

Column (1) shows that the majority of the Germans categorize themselves into health categories three and four. Only four and eight percent are in the worst and best health category, respectively. Columns (3), (5), and (7) show the percentiles of the German,

Canadian, and Finnish sample by SAH category. It is easy to see that the German SF12 is much more symmetric around the value 0.5 and that the other two measures are more skewed towards one. The percentiles are then used as boundaries in an interval regression to scale SAH. Unsurprisingly, the differences in the percentiles translate into differences in the predicted category means. Take category three – good SAH – as an example: while the predicted mean for SF12 is 0.49 (actual mean: 0.48), the predicted mean for the HUI-III cardinalization amounts to 0.82 and the one for 15D amounts to 0.85. A priori, one would expect the differences in the distributions of the three generic health measures to translate into differences in the inequality measures.

Testing Reporting Heterogeneity

The basis of any test on reporting heterogeneity is an objective health measure against which the suspicious variable is tested. One possibility is to use vignettes (Bago d’Uva et al., 2008). However, as correctly stated by Lindeboom and van Doorslaer (2004), it is quite likely that the rating of a health situation hinges crucially on the respondent’s own health status. This, in turn, makes it necessary to condition on the true health status of the respondent, which is not only very data-demanding but also creates a chicken and egg scenario.

Most studies use a generic health measure or mortality as a true health indicator and condition on it together with other covariates within a regression framework (Humphries and Van Doorslaer, 2000; Lindeboom and van Doorslaer, 2004; Etilé and Milcent, 2006). Then, socioeconomic controls, as well as their interactions, should turn out to be insignificant in the absence of differential reporting.¹

Studies that rely on subsequent mortality use this objective health indicator as a dependent variable and condition on SAH arguing that censoring and sample attrition can be dealt with in this way. However, these studies also face the issue that due to the time gap between interview and death, SES-related reporting heterogeneity cannot be unambiguously identified.

¹ Jürges (2008) points out that “the absence of interaction effects is not sufficient to conclude that comparability of SAH across groups is unproblematic.”

Essentially, we will regress various subjective health measures on more objective measures, controlling at the same time for a rich set of socioeconomic controls and their interactions. We estimate various types of models and exploit various subjective, quasi-objective, and objective health measures. We then draw conclusions from the overall pattern of results.

3 Dataset and Variable Definitions

Dataset

The analysis' underlying dataset is the German Socio-Economic Panel Study (SOEP). The SOEP was founded in 1984 and recently sampled more than 20,000 individuals from more than 10,000 households. Wagner et al. (2007) provide more details about this longitudinal household survey. Since wave W (2006) of the SOEP incorporates an extraordinary richness of health measures, we focus on the year 2006 and use all observations without item-non response. The (unweighted) number of respondents total 19,529 for the general specifications, while the grip strength was only taken from a representative sample of 4,311 respondents.

Health Measures

The wordings of the questions that the survey respondents answered can be found in Appendix A.

Subjective Health Measures

Probably the most subjective health measure in the questionnaire is about respondents' health worries. All respondents were asked whether they were concerned about their health. We collapsed the three answer categories into two and generated a dummy variable called *health worries* taking on a zero for those who indicated that they were "not concerned at all" about their health. A one was assigned to those respondents who indicated that they

were “somewhat” or “very concerned.”

A relatively new strand of economics’ literature deals with happiness. The central question upon which this literature builds is about people’s life satisfaction in general – usually measured on a scale from zero to ten. In light of this, it seems odd that little attention has been devoted to measures of health satisfaction. Happiness economists argue that one should rather focus on people’s perceived life satisfaction than on hard economic measures like income per capita. Analogously, one could carry the argument over to the literature of health inequalities and ask whether it would not be more appropriate to focus on people’s health satisfaction rather than their objective health status. This would mean that socio-economic related variation in health perception would be explicitly allowed for and reporting heterogeneity would no longer be an issue .

The SOEP contains a health satisfaction question which is measured on a scale from zero to ten with zero being the lowest value of health satisfaction. We dichotomized the ordinal variable. *Health low* has a one for respondents who indicated values between zero and three.

The most popular health measure is self-assessed health (SAH). Respondents are usually asked how they would rate or describe their health status on a five-category scale, ranging from “bad” to “very good” in the case of the SOEP. Using all five health categories, we estimate generalized ordered logit models. In addition we cardinalize the ordinal SAH measure using the distributions of three generic health measures: SF12, HUI-III, and ED15. Finally, we dichotomize SAH. *Health poor* has a one when the respondents indicated that they were in “poor” or “bad” health. By contrast, *health good* has a one for those who claimed to have a “good” or “very good” health status.

The majority of the literature on health inequalities relies on SAH since it is easy to collect and thus widely available. SAH has been shown to be highly correlated with more objective health measures and is a very reliable indicator of mortality (Appels et al., 1996; Kaplan et al., 1996; McGee et al., 1999). Flaws related to heterogeneous reporting behavior have already been discussed.

Medical Care Measures

Germany has universal health insurance coverage. Around 90 percent of the population are insured under a public health insurance scheme with almost full cost coverage. There are only very modest co-payments which, for instance, amount to €10 per quarter for the outpatient sector.² There are no waiting lists and insured people have free choice of doctor. Conditional on opportunity costs in the form of time lost, the number of doctor visits and the number of days in hospital should be good objective indicators of the respondents' health statuses. The same argument should be true for the number of absence days due to sickness (for those who are employed), given that German social legislation guarantees 100 percent sick pay for up to six weeks per sickness episode.

From the question about the number of doctor visits in the last quarter (see Appendix A), we generate a variable called *doctor visits*. The variable *days in hospital* is also a count variable and has values representing the number of nights in hospital during the previous year. *Dayabs* represents the number of sick leave days taken during the last year for those who were employed.

Quasi-Objective Health Measure: SF12

The SF12 is a generic health measure which was developed to accurately measure the objective health status of respondents. It is comparable to other generic health measures like the HUI-III or the ED15. In total, 12 health-related questions are asked. Using a specific algorithm, two health indices are generated: the first is called *mcs* and measures mental health, whereas *pcs* measures physical health. We also combine both measures by simple weighting and call the overall measure *sf12index*.

As can be seen in Figure 2, the distribution of all three continuous measures is slightly skewed but all three measures show a substantial degree of variation. We rescale all measures to lie between zero and one. Since we measure ill-health in some models, we also generate variables by taking one minus the value of the index and call them *neg_mcs*,

² There are various exemptions from the general rule and people can apply for a case of hardship which rules out co-payments. Thus, low wages earners are usually exempted from co-payments.

neg_pcs, and *neg_sf12index*.

[Insert Figure 2 about here]

More details on how the SF12 was calculated and which questions were asked, can be found in Appendix A.

Objective Health Measure: Grip Strength

The grip strength is an objective measure of physical health. It has been shown to have a highly predictive power with respect to many health outcome measures (Giampaoli et al., 1999; Rantanen et al., 1999, 2000; Metter et al., 2002). The grip strength is non-invasive, easy to measure and hence the response rate in the SOEP lies above 96 percent. Selected interviewers were trained to accurately measure the grip strength of the respondents. More details about the measurement of the grip strength can also be found in Appendix A.

As with SF12, Figure 2 shows a large degree of variation in the distribution of the grip strength (bottom right graph). And as before, we rescale the grip strength to the zero-one interval and subtract the index from one to get a measure of ill health.

Explanatory Variables

To standardize the health measure with respect to age and gender, we generate ten gender-specific age groups (<35 , $35-44$, $44-64$, $65-74$, >75) and call them *mage1-mage5* and *fage1-fage5*. Our main SES measure is the equalized household income, which we generate according to the following formula:

$$EHI_i = \frac{HHI_j}{\omega_j}$$

where EHI_i stands for the equalized household income of person i and is the ratio of the monthly net income of household j , HHI_j , and the household weight ω_j , which is computed according to:

$$\omega_j = 1 + (m_j - 1 - k_j) \times 0.5 + (k_j \times 0.3)$$

with m_j being the total number of household members and k_j being the number of children under the age of 14 in the household. The latter are weighted by 0.3 as compared to members over 14, which are weighted by a factor of 0.5.

To measure the educational level, we generate four dummy variables, namely *dropout* as well as *9*, *10*, and *13 years of schooling*, respectively. In addition, we have a dummy *apprenticeship* for those who hold such a degree and a dummy *college* for those with a college degree. We also make use of the following socio-economic dummy variables: *full-time employed*, *kids*, *immigrant*, *married*, *weight*, and *height*. Besides interacting each variable with the income measure, we also employ a dummy *East Germany*.

4 Results

We start by looking at the mean values of all health measures. They are displayed in Table 2 by six income percentile categories: <p5, p5-p25, p25-p50, p50-p75, p75-p95, >p95. Besides displaying the pure mean values, we also display the weighted means as well as gender-age standardized and weighted means.³

[Insert Table 2 about here]

The first finding is that the means of all subjective ill-health measures drop substantially, the higher the income category. Analogously, we find a continuous increase in health – though less pronounced – by income category when we look at the scaled SAH measures. As for *days in hospital*, we also find a distinctive decrease by income class, while the number of doctor visits and sick leave days peak in the second and third lowest income category and then decline monotonically. Looking at the mental health index of the SF12, we find

³ The gender-age standardization has been performed according to the indirect standardization method as described in O'Donnell et al. (2007).

monotonically increasing health states by income categories. The evolution of the SF12 physical health component is much flatter and we find the lowest value for the second lowest income class. This resembles what we find for the grip strength: the values do not vary much between income groups and increase from the second category to the fifth, having the lowest and highest value in each of these categories.

A second finding is that weighting and standardizing the health measures change their mean values, but only very slightly. The stars in the table indicate significant differences between the raw mean values and the weighted and standardized values. Significant differences are consistently found for all health measures but seem to be particularly important for the extreme income classes. The size of the differences is always relatively small.

To visualize the association between income and health, we plot cumulative density functions (cdfs) of SF12 measures and grip strength by income category. In Figure 3, we observe a clear and distinct health-income gradient for the mental health index as well as for the overall index of the SF12. However, the positive association between income and health is less pronounced for the physical component of the SF12 and barely visible for the grip strength.

[Insert Figure 3 about here]

In a further step, we look at the concentration curves by category of health measures. Figure 4 plots the curves for the dichotomized subjective health measures. We see a distinct degree of health inequality with the graphs for poor health lying substantially above the degree of equality and the graph for good health lying below that line.

[Insert Figure 4 about here]

The concentration curves for measures that indicate the degree of usage of the health care system are plotted in Figure 5. While the curve for *days in hospital* resembles the curves in Figure 4 and displays a distinct degree of health inequality with a curve lying constantly above the 45°-line, the other two graphs differ. At first glance, it is difficult to detect any difference between the 45°-line and the concentration curve for the number of

doctor visits. In a higher resolution and magnified, one can see that the curve crosses the line of equality several times up to the value of 0.5 on the abscissa and lies above it for higher values. The number of sick leave days shows a similar pattern. From value 0.4 onwards, sick leave behavior is more concentrated among the poor, whereas the concentration curve lies below the 45°-line for smaller values – a finding that is in line with the literature (Wagstaff and van Doorslaer, 2000).

[Insert Figure 5 about here]

Figure 6 uses one minus the corresponding SF12 or grip strength measure as an ill-health indicator. Consequently, one would expect the *CCs* to lie above the 45°-lines in the case of pro-poor health inequalities. What we find is that the two lines are not apparently distinguishable. However, enlarging the picture reveals that all concentration curves lie slightly above the lines of equality. Inequality is most pronounced for the grip strength.

[Insert Figure 6 about here]

We have tested all concentration curves for dominance against each other (see Appendix B).⁴ In general, the curves of the subjective health measures dominate all other curves. We only find non-dominance when they are tested against *days in hospital* and *daysabs* since the pro-poor deviation of the latter two curves is very large for the richer half of the sample. The scaled SAH curves (not displayed) mostly dominate the medical care measures. As expected, the curve of the SF12 measure and its SAH scaled variant cross. Since the German SF12 has less unequal health distribution than the HUI-III and the 15D, the scaled SAH version of the SF12 is dominated by the other two variants. As we have seen in Figure 6, the *CCs* of the SF12 and the grip strength are very close to the 45°-line; thus, they are almost always dominated by the other curves.

Table 3 shows concentration indices of all health measures. All indices differ significantly from zero and have negative signs, indicating that ill-health is more concentrated

⁴ Dominance testing was performed by means of the user-written Stata program `dominance` which can be downloaded from <https://www.worldbank.org/analyzinghealthequity>.

among the poor. However, there are large differences in size, depending on the underlying health measure. We also find differences, depending on whether the health measure is age-gender standardized or not and whether education or income is used as socioeconomic indicator. Interestingly, these differences in size are negligible as compared to the differences between the groups of health measures used.

[Insert Table 3 about here]

The largest degree of inequality is found when dichotomized subjective health measures are the basis of the calculation, with values ranging from -0.1 to -0.15. Inequality decreases when medical care indicators are employed. Concentration indices for this type of health measure range between -0.05 and -0.1. The number of doctor visits is one of the few exceptions where the choice of the underlying SES measure has an impact. Related to income, the indices tend towards zero and are not significantly different from zero.

Using the SF12 as a health measure, by component or as a whole, reduces the degree of health inequality dramatically. It oscillates around -0.01 which is a ten times smaller than the dichotomized subjective indicators. We get the same value when we use grip strength as a measure of physical health. Notice that the standardized *CI*s for the SF12 physical health component and grip strength are the same size – no matter whether we use income or education as an SES measure. It is also noteworthy that the degree of inequality for the SF12 mental health component is substantially smaller when respondents are ranked by their education rather than by their household income. This is in line with the notion that it is economically precarious situations rather than poor education that causes psychological stress.

Once we scale SAH with SF12, the size of the resulting *CI* is very similar to the *CI* of the SF12 rather than the *CI* of the dichotomized SAH measure. This finding shows that the distribution of a generic health measure determines the degree of inequality when SAH is scaled by means of this generic health measure. The consideration is reinforced when the *CI*s of the other scaling procedures are looked at. Scaling SAH by means of the Canadian HUI-III leaves us with a *CI* of about -0.06, whereas it amounts to -0.05 when the Finish 15D is used. The values are very similar to those reported by Lauridsen et al. (2004) who

compared the HUI-III to the 15D as a scaling instrument using Finnish data.⁵ This exercise shows that the measured degree of inequality is highly sensitive to the underlying scaling instrument. Hence, researchers should routinely report robustness checks with alternative underlying scaling instruments when scaled SAH is used to calculate health inequality indices.

Table 5 (Appendix B) shows the results when mean equality of all 13 *CI*s is formally tested. Almost all *CI*s differ significantly from each other; particularly in comparing different groups of health measures, we find highly significant differences in means. Again, one exception is *doctor visits* whose *CI* is close to zero and not statistically different from the SF12 based *CI*s.

In Table 4 we provide regression estimates of the different health measures on more objective health measures and various controls. In columns (1) to (8), we regress the dichotomized subjective measures and SAH on the preference-based SF12 which is claimed to provide a comprehensive and unbiased measure of health. If this is true, other socio-economic characteristics should not be significantly related to these health measures once the objective health status is controlled for.

[Insert Table 4 about here]

As we can easily infer from the table, SF12 is by far the most important and a highly significant determinant of subjective health measures. As additional controls, we generated eight gender-age dummies which entered the models in levels and in interaction with income. To save space, their coefficients are not shown. For each model with dichotomized subjective health (columns (1) to (3)), we find four gender-age dummies and their income interactions to be significant. If we turn to the generalized ordered logit and SAH, we find that few of the gender-age dummies are significant for the lower health categories. However, for the SAH-3 and SAH-4 categories, all gender-age dummies are significant. Moreover, the coefficients differ in size.⁶ Regarding their interactions with income, we find

⁵ In contrast to our study, Lauridsen et al. (2004) use positive health measures and find *CI*s of 0.01 and 0.015 for SAH cardinalized by the 15D and the HUI-III. If we define our scaled SAH in the same way, the result is *CI*s of 0.007 and 0.012, respectively.

⁶ Separate statistical testing reveals that *mage2-5* and *fage2-5* are jointly statistically different from zero. By SAH categories, various coefficients are statistically different from each other in their means.

three out of eight coefficients to be significant for the categories SAH-3 to SAH-5. Again, some of these coefficients are statistically different in their means.

While gender-age specific reporting heterogeneity has been consistently found in the literature, the question of whether education or income-related reporting heterogeneity is an issue is more controversial. Besides the fact that we have found various gender-age-income interactions to be significant, the table provides additional evidence for SES-related reporting differentials. In some of the regression models in columns (1) to (8), income in levels and in its squared form is correlated with the reporting behavior. We also see that various educational covariates, in levels and in interaction with income, show a significant association with subjective health measures, conditional on a presumably objective and comprehensive health measure. Interestingly, except for the model with low health satisfaction as dependent variable, we also observe that interactions between income and the employment status, the marital status and *kids* are significantly related to subjective health measures. Being full-time employed is clearly negatively associated with ill-health.⁷

Let us turn to the model in column (9) in detail. To obtain evidence for reporting heterogeneity in preference-based generic health measures, we regress the physical health component of the SF12 on an unambiguously objective physical health measure – the grip strength – and a set of controls. Unsurprisingly, grip strength is clearly the most important determinant of SF12 physical health. It turns out that almost all gender-age controls are significant, which is not unexpected since physical health clearly deteriorates with rising age. On the other hand, this finding suggests that a certain degree of gender-age specific reporting heterogeneity is also inherent in generic health indicators. This, in turn, would mean that it is advisable to gender-age standardize such measures before using them as objective controls in tests for or against reporting heterogeneity.

Column (9) shows the importance of considering height and weight when physical health measures are dealt with. The two measures do not show a significant relationship with

⁷ As for the generalized logit and SAH, *full-time employed*, *married*, and *kids* are each jointly statistically different from zero. Moreover, the significant coefficients in the according categories mostly differ in their means from the other coefficients.

the dependent variable in any of the other models. On the other hand, and in contrast to the other subjective health measures, the physical health measure of the SF12 seems to be free of an education-related reporting bias.

This also holds for labor market and marital status related biases. However, it is striking that the regression model in column (9) is the only model that shows a strong and significant association between household income levels and physical health after having controlled for an objective measure of physical health and a rich set of socioeconomic controls. The positive association between higher household income and reported better physical health is as strong as the impact of the grip strength on SF12 physical health. However, the finding is no definite proof for the existence of income-related reporting heterogeneity. Is it imaginable that there exists an unobserved dimension of physical health which is not captured by grip strength or any of the other socioeconomic covariates while being captured by the SF12 physical health measure. If income had the same strong impact on this physical health component as grip strength has on physical health, this strong positive correlation between income and reported physical health could be explained. We believe that this is highly implausible. It is at least implausible that such a relationship accounts for the entire correlation. The more so as Germany has a very generous universal health coverage with low access barriers, short waiting times, and free choice of doctors and hospitals. Low wage earners are publicly insured and mostly exempted from co-payments. Ninety percent of the population have a public health insurance with very moderate co-payments while coinsurance rates and deductibles are banned by German social legislation. Hence it is difficult to think of a story that could explain this substantial correlation but excludes reporting heterogeneity. Consequently, we interpret the very strong linkage between lower income and reported worse physical health net of grip strength and various socioeconomic controls as evidence for income-related reporting heterogeneity in generic health measures.

5 Discussion and Conclusion

This paper systematically assesses the sensitivity of the concentration index (CI) with respect to the underlying health measure. The magnitude of the concentration index varies far more between different categories of health measures than between raw, weighted, or age-gender standardized health measures as well as other living standard measures. Using dichotomized subjective health measures like SAH or health satisfaction yields the largest pro-poor health inequalities. The indices are approximately halved when measures of health care use such as doctor visits are used as health indicators. When using the SF12, a preference-based and widely accepted generic health measure, the CI drops by a factor of ten. However, the CI 's size does not vary according to whether the physical, the mental, or the combined component of the SF12 is used.

Scaling SAH by means of the SF12 and interval regressions results in a similar degree of inequality to the pure SF12. Scaling SAH by means of other generic health measures derived from other populations alters the results significantly. Given that the alternative HUI-III and 15D measures are more than 10 years old and rest upon the Canadian and Finnish population, we propose the usage of the German SF12 when European SAH distributions are scaled. Germany is the most populous European country, the SF12 is taken from a large representative 25-year-old household survey, and the SF12 is carried out at regular intervals. Moreover, the distribution of the SF12 is relatively symmetric and includes an explicit mental health component; the HUI-III has been criticized for its focus on the physical health dimension (Richardson and Zumbo, 2000). However, we suggest presenting routinely scaled SAH for various underlying generic health distributions to prove the robustness of the results.

Finally, we present evidence on reporting heterogeneity. For all health measures, we find strong evidence for gender-age related reporting heterogeneity. The finding is in line with the literature and suggests that health measures should be routinely age-gender standardized. Regarding subjective health indicators, we find additional evidence for socioeconomic-related reporting heterogeneity, depending on the educational level, the labor market status and the marital status of the respondents. The latter forms of differential reporting are

not inherent in the physical component of the SF12, which, in contrast, shows clear associations with age, gender, height, and weight. Moreover, we find a strikingly strong and positive correlation between income and physical health in the model where we regress the physical health component of the SF12 on grip strength as an objective physical health measure along with a rich set of other covariates. In this way, we provide evidence on the presence of reporting heterogeneity in preference-based generic health measures.

In accordance with claims made by advocates of the economics of happiness, measures of health inequality could be routinely based on health satisfaction rather than, or in addition to, presumably objective health measures. Socioeconomic-related reporting behavior would be explicitly allowed for and, hence, reporting heterogeneity would no longer be a flaw. In addition, measures of health satisfaction can be as easily collected as SAH and they rely on an increasingly popular strand in the economics literature. In this way, the fundamental identification problem to disentangle objective health from differential reporting could be overcome.

References

- Andersen, H. H., A. Mühlbacher, M. Nübling, J. Schupp, and G. G. Wagner (2007). Computation of standard values for physical and mental health scale scores using the SOEP version of sf12v2. *Journal of Applied Social Science Studies (Zeitschrift für Wirtschafts- und Sozialwissenschaften)* 127(1), 171–182.
- Appels, A., H. Bosma, V. Grabauskas, A. Gostautas, and F. Sturmans (1996). Self-rated health and mortality in a Lithuanian and a Dutch population. *Social Science & Medicine* 42(5), 681–689.
- Bago d’Uva, T., O. O’Donnell, and E. van Doorslaer (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology* 37(6), 1375–1383.
- Baker, M., M. Stabile, and C. Deri (2004). What do self-reported, objective, measures of health measure? *Journal of Human Resources* 39(4), 1067–1093.

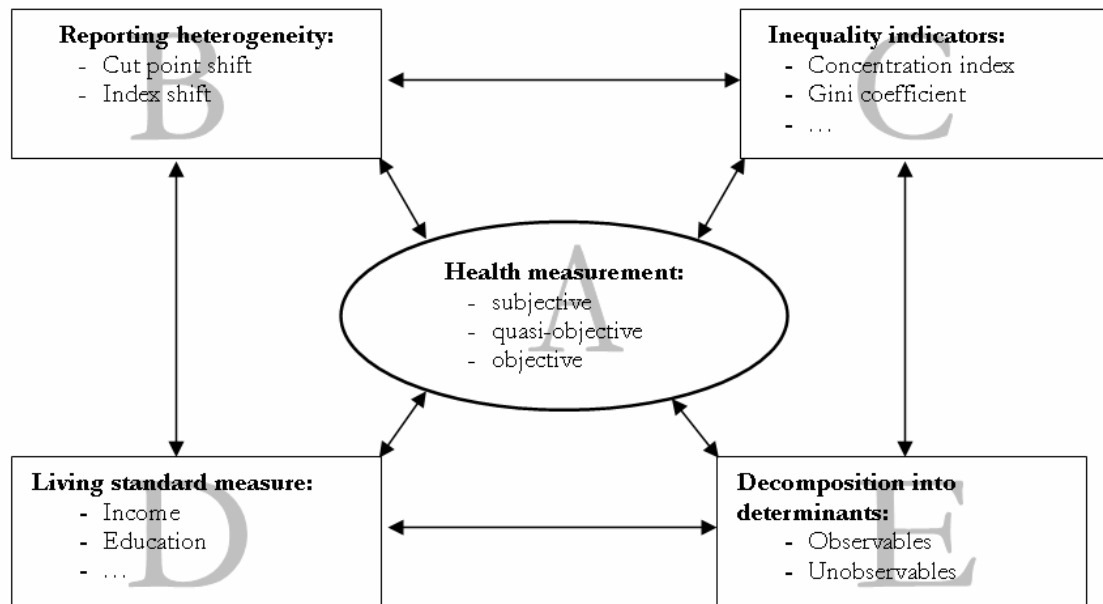
- Erreygers, G. (2009). Correcting the concentration index. *Journal of Health Economics* 28(2), 504–515.
- Etilé, F. and C. Milcent (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from france. *Health Economics* 15(9), 965–981.
- Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2005). The causal effect of income on health: Evidence from German reunification. *Journal of Health Economics* 24(5), 997–1017.
- Giampaoli, S., L. Ferrucci, F. Cecchi, C. Lo Noce, A. Poce, F. Dima, A. Santaquilani, M. F. Vescio, and A. Menotti (1999). Hand-grip strength predicts incident disability in non-disabled older men. *Age and Ageing* 28(3), 283–288.
- Hank, K., H. Jürges, and G. G. Wagner (2009). Isometrische Greifkraft und sozialgerontologische Forschung: Ergebnisse und Analysepotentiale des SHARE und SOEP. *Zeitschrift für Gerontologie und Geriatrie* 42(2), 117–126.
- Huber, H. (2008). Decomposing the causes of inequalities in health care use: A micro-simulations approach. *Journal of Health Economics* 27(6), 1605–1613.
- Humphries, K. H. and E. Van Doorslaer (2000). Income-related health inequality in Canada. *Social Science & Medicine* 50(5), 663–671.
- Johnston, D. W., C. Propper, and M. A. Shields (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics* 28(3), 540–552.
- Jones, A. M., N. Rice, T. B. d’Uva, and S. Balia (2007). *Applied Health Economics* (1 ed.). Routledge.
- Jürges, H. (2007). True health vs. response styles : Exploring cross-country differences in self-reported health. *Health Economics* 16(2), 163–178.
- Jürges, H. (2008). Self-assessed health, reference levels and mortality. *Applied Economics* 40(5), 569–582.

- Kakwani, N., A. Wagstaff, and E. van Doorslaer (1997). Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics* 77(1), 87–103.
- Kaplan, G. A., D. E. Goldberg, S. A. Everson, R. D. Cohen, R. Salonen, J. Tuomilehto, and J. Salonen (1996). Perceived health status and morbidity and mortality: Evidence from the Kuopio Ischaemic Heart Disease Risk Factor Study. *International Journal of Epidemiology* 25(2), 259–265.
- Lauridsen, J., T. Christiansen, J. Gundgaard, U. Häkkinen, and H. Sintonen (2007). Decomposition of health inequality by determinants and dimensions. *Health Economics* 16(1), 97–102.
- Lauridsen, J., T. Christiansen, and U. Häkkinen (2004). Measuring inequality in self-reported health—Discussion of a recently suggested approach using Finnish data. *Health Economics* 13(7), 725–732.
- Lindeboom, M. and E. van Doorslaer (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics* 23(6), 1083–1099.
- Lindelow, M. (2006). Sometimes more equal than others: How health inequalities depend on the choice of welfare indicator. *Health Economics* 15(3), 263–279.
- Mathers, C. D. and R. Douglas (1998). Measuring progress in population health and well-being. In R. Eckersley (Ed.), *Measuring progress: Is life getting better*, pp. 125–155. Collingwood, Ontario, Canada: CSIRO Publishing.
- McGee, D. L., Y. Liao, G. Cao, and R. S. Cooper (1999). Self-reported health status and mortality in a multiethnic US cohort. *American Journal of Epidemiology* 149(1), 41–46.
- Metter, E. J., L. A. Talbot, M. Schrager, and R. Conwit (2002). Skeletal muscle strength as a predictor of all-cause mortality in healthy men. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* A57(10), B359–B365.

- O'Donnell, O., E. van Doorslaer, A. Wagstaff, and M. Lindelow (2007). *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and their Implementation* (1 ed.). World Bank Publications.
- Rantanen, T., J. M. Guralnik, D. Foley, K. Masaki, S. Leveille, J. D. Curb, and L. White (1999). Midlife hand grip strength as a predictor of old age disability. *Journal of the American Medical Association* 281(6), 558–560.
- Rantanen, T., T. Harris, S. G. Leveille, M. Visser, D. Foley, K. Masaki, and J. M. Guralnik (2000). Muscle strength and body mass index as long-term predictors of mortality in initially healthy men. *Journal of Gerontology: Medical Sciences* 55A(3), M168–M173.
- Richardson, C. G. and B. D. Zumbo (2000). A statistical examination of the health utility index-mark iii as a summary measure of health status for a general population health survey. *Social Indicators Research* 51(21), 171–191.
- Schupp, J. (2007). Greifkraftmessung im Sozio-oekonomischen Panel (SOEP). Data Documentation 23, DIW Berlin. [http : //www.diw.de/documents/publikationen/73/60179/diw_datadoc2007 - 023.pdf](http://www.diw.de/documents/publikationen/73/60179/diw_datadoc2007-023.pdf), last accessed at May 15, 2009.
- Tubeuf, S. (2008). Income-related inequalities in self-assessed health: Comparisons of alternative measurements of health. Health, Econometrics and Data Group (HEDG) Working Papers, University of York.
- van Doorslaer, E. and U.-G. Gerdtham (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from swedish data. *Social Science & Medicine* 57(9), 1621–1629(9).
- van Doorslaer, E. and A. M. Jones (2003). Inequalities in self-reported health: Validation of a new approach to measurement. *Journal of Health Economics* 22(1), 61–87.
- van Doorslaer, E. and X. Koolman (2004). Explaining the differences in income-related health inequalities across European countries. *Health Economics* 13(7), 609–628.

- Wagner, G. G., J. R. Frick, and J. Schupp (2007). The German Socio-Economic Panel study (SOEP)—Evolution, scope and enhancements. *Journal of Applied Social Science Studies (Schmollers Jahrbuch)* 127(1), 139–169.
- Wagstaff, A. (2005). The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics* 14(4), 429–432.
- Wagstaff, A. and E. van Doorslaer (2000). Equity in health care finance and delivery. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1, Chapter 34, pp. 1803–1862. Elsevier.
- Wagstaff, A. and N. Watanabe (2003). What difference does the choice of ses make in health inequality measurement? *Health Economics* 12(10), 885–890.

Figure 1: Classification of the Research Field of Health Inequalities



Source: own illustration

Figure 2: Pdfs of Generic Health Measures (SF12) and Grip Strength

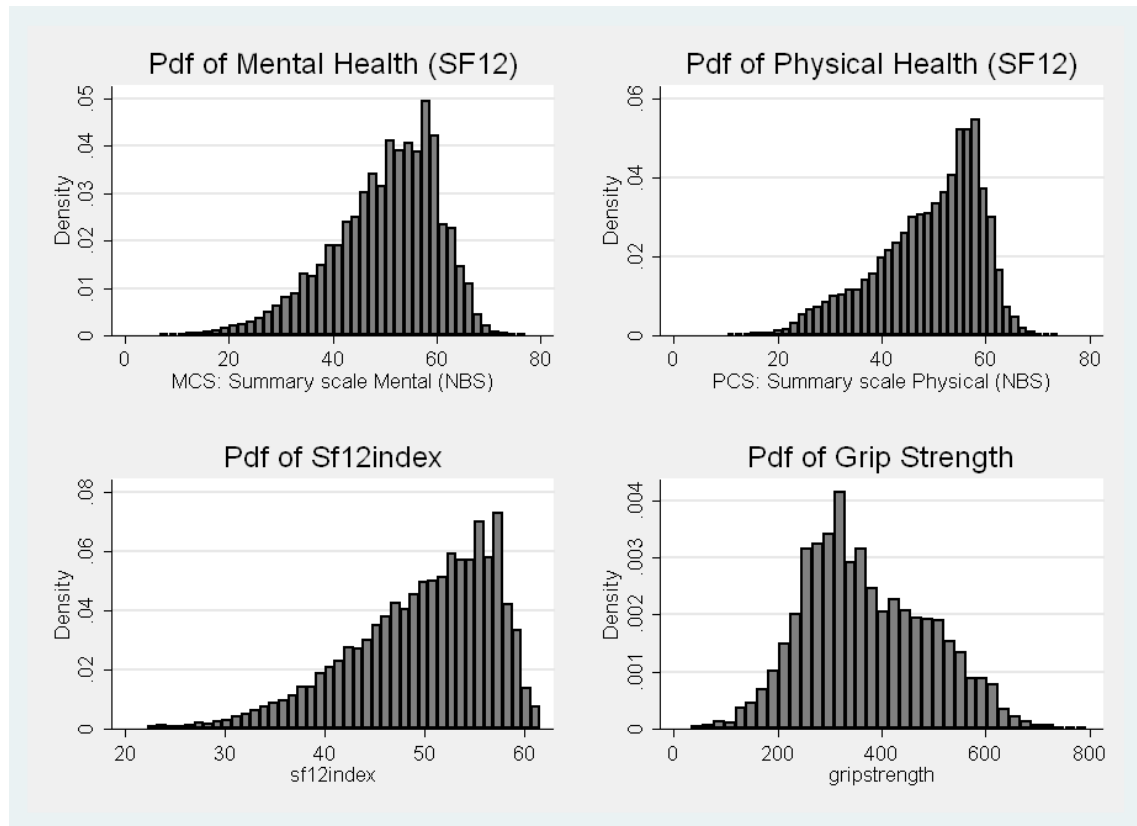


Figure 3: Cdfs of Generic Health Measures (SF12) and Grip Strength by Income Category

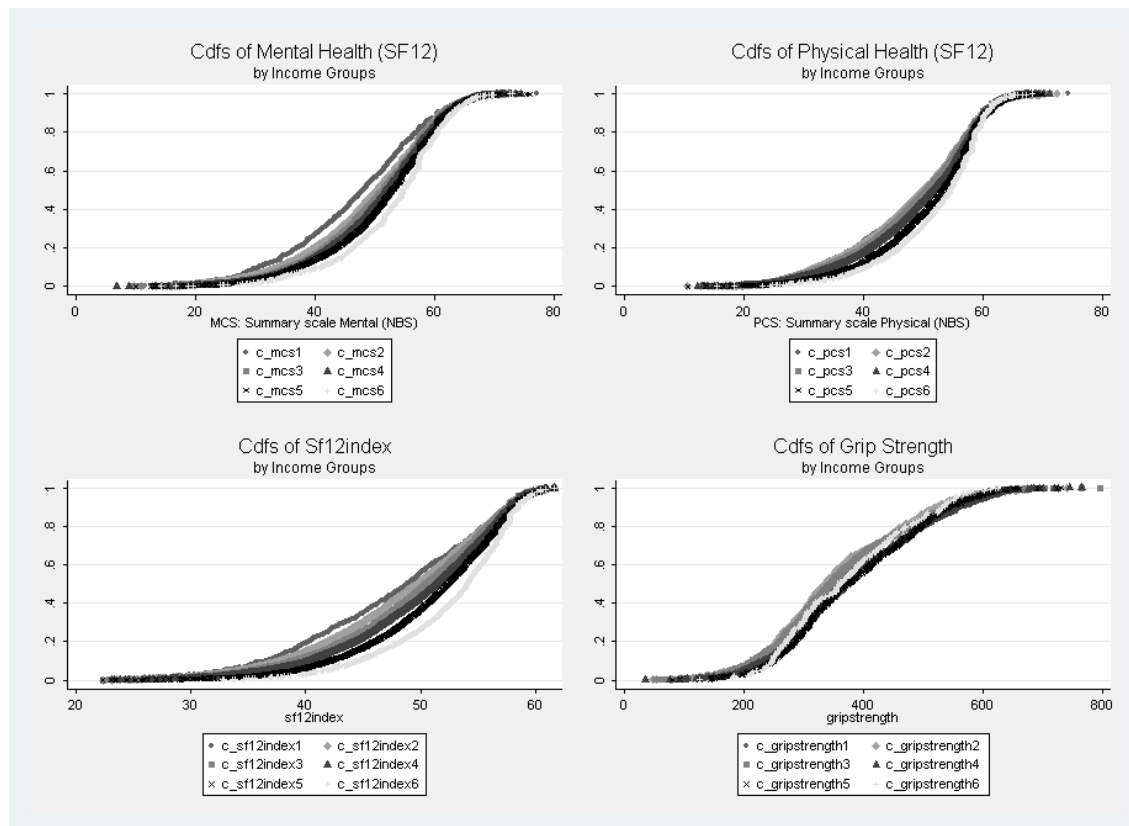


Figure 4: Concentration Curves of Subjective Health Measures

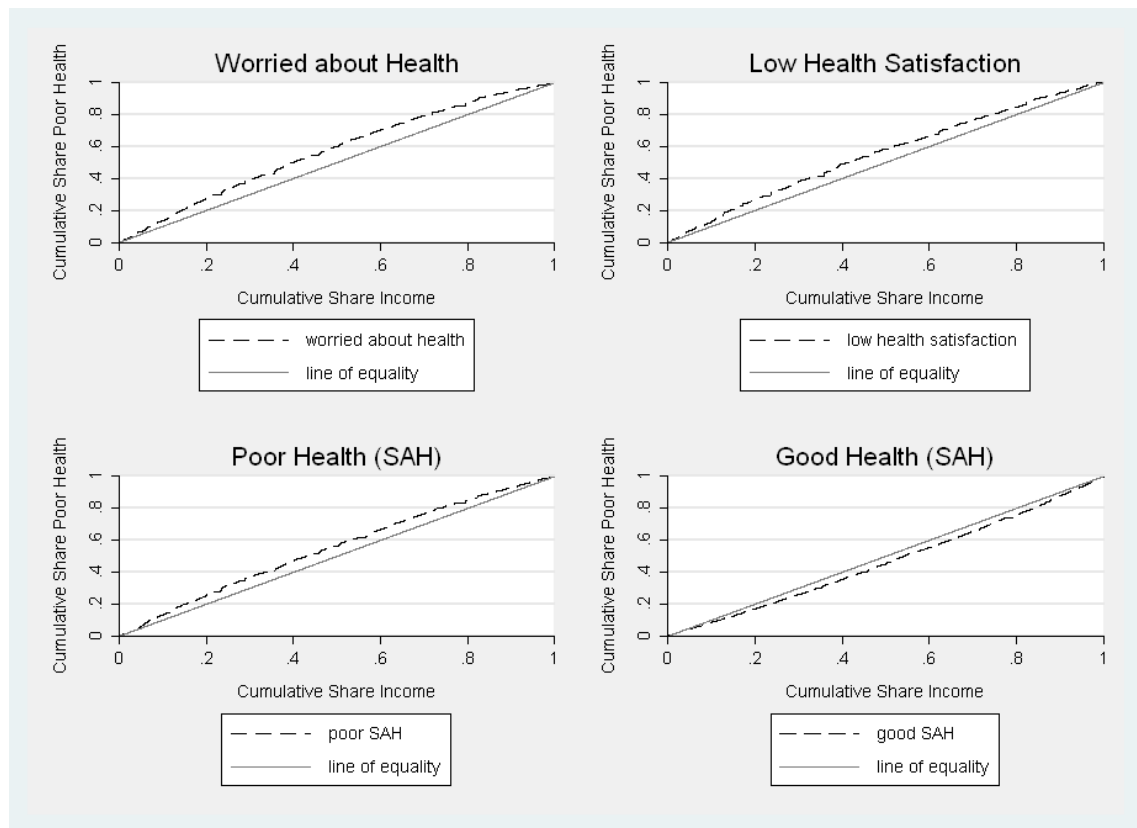


Figure 5: Concentration Curves of Health Care System Use Measures

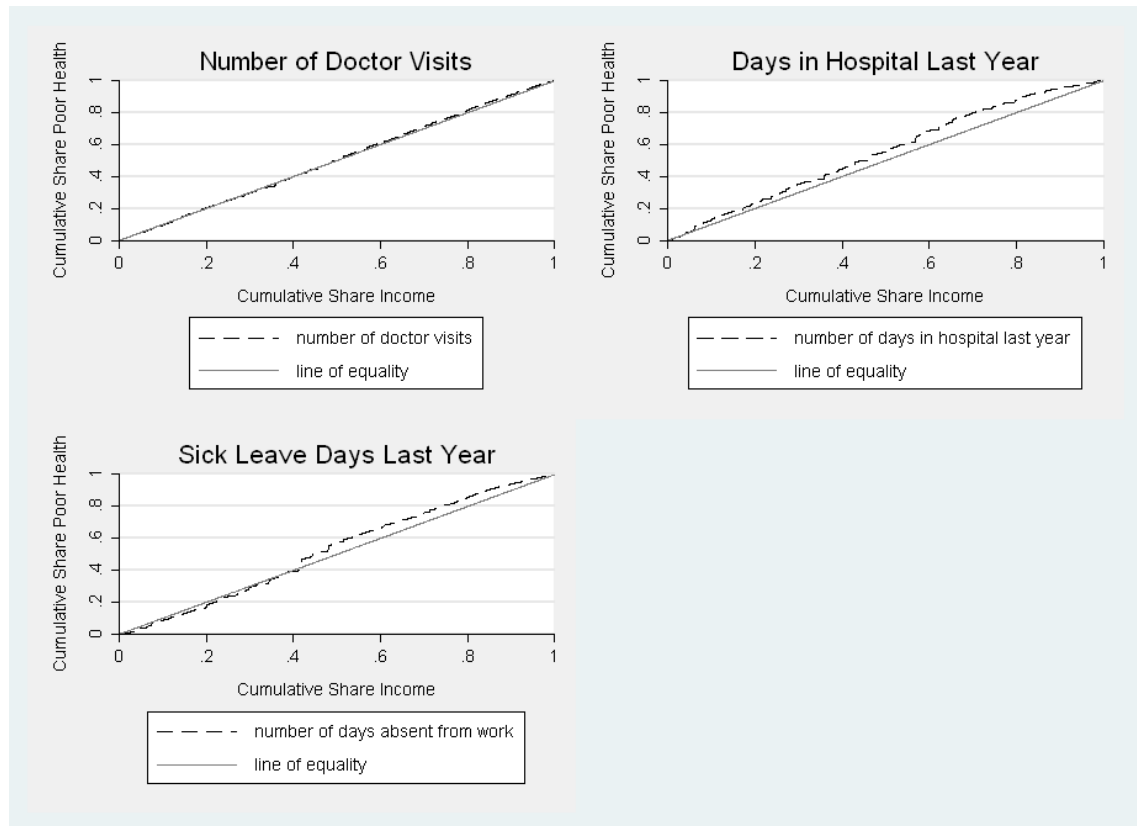


Figure 6: Concentration Curves of Generic Health Measures (SF12) and Grip Strength

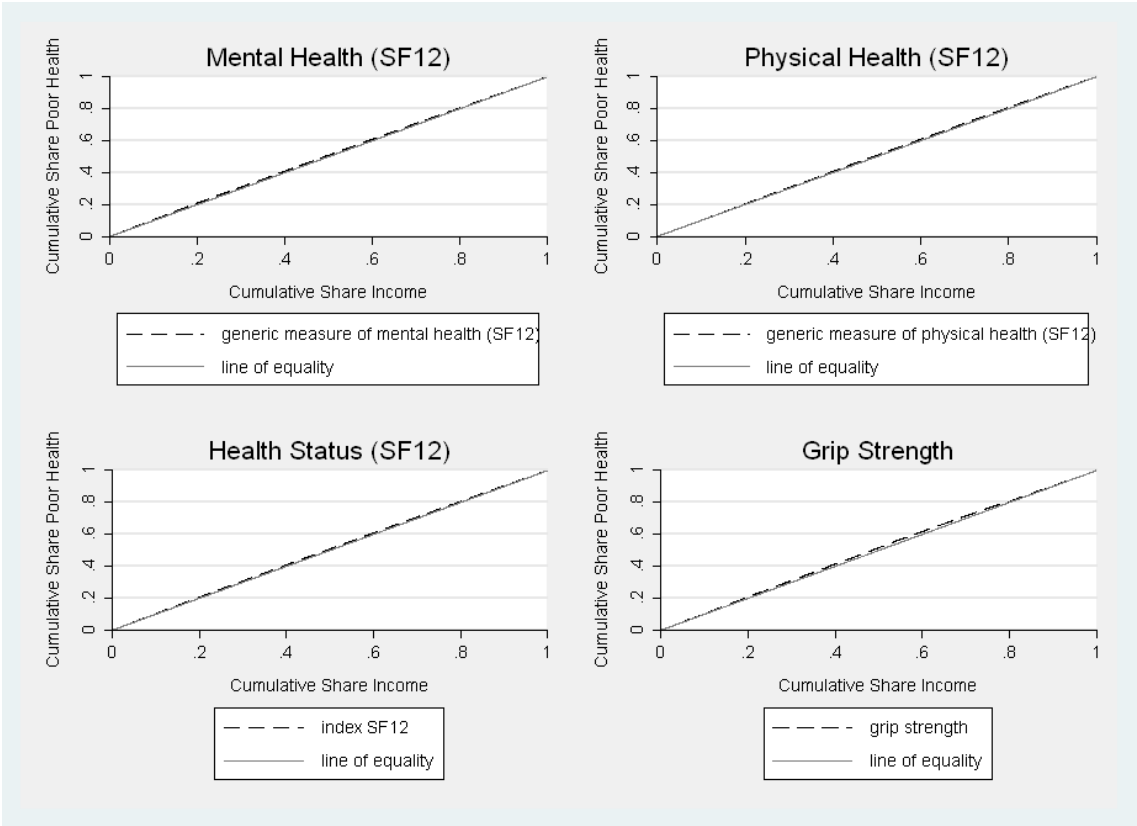


Table 1: Summary Statistics of SF12, HUI-III, and 15D by SAH categories

SAH	N (%)	SF12		HUI-III		15D	
		Mean	Percentile	Mean	Percentile	Mean	Percentile
1: Poor	785 (0.040)	0.3506	0.3374	0.3694	0.4119	0.6408	0.673
2: Fair	2,673 (0.137)	0.4214	0.4184	0.6539	0.746	0.7478	0.789
3: Good	6,523 (0.334)	0.4932	0.5063	0.8208	0.897	0.8549	0.917
4: Very Good	7,934 (0.406)	0.5551	0.6776	0.9211	0.947	0.9395	0.965
5: Excellent	1,614 (0.083)	0.5977	1	0.9721	1	0.9813	1

Source: German Socio-Economic Panel Study (SOEP)

Values for HUI-III are taken from Jones et al. (2007); values for 15D are taken from Lauridsen et al. (2004)

In the SOEP, the translation of the SAH categories would read “bad, poor, satisfactory, good, very good.”

In the Finnish Health Care Survey, from which the 15D values are taken, the translation of the categories would read “poor, rather poor, average, fairly good, good.”

The “mean”-columns show the predicted values (conditional on the category) when the according generic health measure is used to scale SAH. The procedure is applied like in (Jones et al., 2007). Interval regressions are weighted.

The actual mean values of the SF12 by SAH category are: 0.3364, 0.4107, 0.4799, 0.5342, 0.5713

Table 2: Health Measures by Categories of Equivalized Household Income per Respondent

Health Measure	< p5			p5 < Income < p25			p25 < Income < p50		
	raw	wght.	wght. & std.	raw	wght.	wght. & std.	raw	wght.	wght. & std.
Subjective measures									
Health worries	0.284	0.288	0.299***	0.247	0.257	0.260***	0.209	0.214	0.210
Low health satisfaction	0.158	0.181	0.188***	0.131	0.148	0.153***	0.114	0.122	0.121
High health satisfaction	0.372	0.342	0.320***	0.375	0.361	0.351***	0.394	0.380	0.381
Poor SAH	0.249	0.268	0.282***	0.222	0.238	0.242***	0.188	0.196	0.193
Good SAH	0.430	0.402	0.375***	0.437	0.395	0.386***	0.466	0.452	0.456
SAH scaled by generic measures (<i>interval regression</i>)									
SF12	0.486	0.482	0.478***	0.496	0.489	0.489**	0.503	0.500	0.501
HUI-III	0.807	0.800	0.792***	0.815	0.804	0.801***	0.826	0.820	0.822
15D	0.861	0.856	0.851***	0.865	0.858	0.856***	0.872	0.868	0.869
Use of health services									
Doctor visits	2.463	2.464	2.542***	2.510	2.511	2.508	2.548	2.648	2.617**
Days in hospital	1.973	2.354	2.443***	1.835	1.830	1.861**	1.636	1.704	1.649
Sick leave days	7.804	8.428	7.971***	8.959	8.739	8.154***	8.385	8.797	8.146***
Generic health measure (SF12)									
Mental health	0.469	0.462	0.464***	0.489	0.479	0.480	0.499	0.495	0.494***
Physical health	0.486	0.482	0.475***	0.481	0.476	0.475***	0.486	0.483	0.485
Combined index	0.478	0.472	0.469***	0.485	0.478	0.477***	0.493	0.489	0.490
Grip strength	0.443	0.407	0.405***	0.422	0.408	0.416**	0.431	0.413	0.416**

Continued on next page...

... Table 2 continued

Health Measure	p50 < Income < p75			p75 < Income < p95			>p95		
	raw	wght.	wght. & std.	raw	wght.	wght. & std.	raw	wght.	wght. & std.
Subjective measures									
Health worries	0.167	0.174	0.175**	0.104	0.134	0.131	0.080	0.088	0.087***
Low health satisfaction	0.096	0.109	0.109	0.077	0.104	0.099**	0.051	0.037	0.032***
High health satisfaction	0.422	0.398	0.398	0.488	0.464	0.479***	0.553	0.508	0.523***
Poor SAH	0.170	0.183	0.183*	0.129	0.155	0.149	0.090	0.098	0.093***
Good SAH	0.488	0.459	0.459	0.555	0.536	0.550***	0.621	0.578	0.592***
SAH scaled by generic measures (<i>interval regression</i>)									
SF12	0.512	0.506	0.505***	0.526	0.519	0.520	0.545	0.538	0.539***
HUI-III	0.835	0.828	0.828	0.854	0.846	0.850*	0.875	0.869	0.873***
15D	0.877	0.873	0.873	0.891	0.885	0.888**	0.904	0.899	0.901***
Use of health services									
Doctor visits	2.430	2.570	2.588***	2.394	2.446	2.438***	1.939	1.794	1.817***
Days in hospital	1.593	1.740	1.754***	1.029	1.047	1.027***	0.631	0.711	0.725**
Sick leave days	7.803	8.722	7.867***	6.626	7.082	5.937***	4.396	4.139	2.959***
Generic health measure (SF12)									
Mental health	0.506	0.497	0.497***	0.513	0.506	0.507***	0.531	0.516	0.515**
Physical health	0.494	0.491	0.490**	0.510	0.503	0.505	0.522	0.523	0.524***
Combined index	0.500	0.494	0.493***	0.511	0.505	0.506	0.527	0.519	0.520***
Grip strength (rescaled)	0.459	0.454	0.439***	0.462	0.455	0.438***	0.444	0.422	0.426

* p<0.1, ** p<0.05, *** p<0.01 indicating differences in means between the raw and the standardized measure

p stands for percentiles

std. stands for an gender-age indirect standardization

indirect gender-age standardization was performed according to O'Donnell et al. (2007)

Table 3: Concentration Indices of Health Measures

Health measures	Income		Education	
	wght. (1)	wght. & std. (2)	wght. (3)	wght. & std. (4)
Subjective measures				
Health worries	-0.145*** (0.008)	-0.151*** (0.008)	-0.130*** (0.007)	-0.120*** (0.007)
Low health satisfaction	-0.114*** (0.010)	-0.131*** (0.010)	-0.112*** (0.009)	-0.112*** (0.009)
Poor SAH	-0.103*** (0.008)	-0.115*** (0.008)	-0.115*** (0.007)	-0.105*** (0.007)
SAH scaled by generic measures				
SF12	-0.014*** (0.001)	-0.015*** (0.001)	-0.014*** (0.001)	-0.015*** (0.001)
HUI-III	-0.055*** (0.003)	-0.063*** (0.003)	-0.056*** (0.003)	-0.064*** (0.003)
15D	-0.047*** (0.003)	-0.054*** (0.003)	-0.048*** (0.002)	-0.055*** (0.002)
Use of health services				
Doctor visits	-0.009 (0.006)	-0.010 (0.006)	-0.054*** (0.006)	-0.046*** (0.005)
Days in hospital	-0.105*** (0.017)	-0.110*** (0.017)	-0.083*** (0.015)	-0.059*** (0.014)
Sick leave days	-0.064*** (0.007)	-0.088*** (0.008)	-0.091*** (0.012)	-0.112*** (0.014)
Generic health measure (SF12)				
Mental health	-0.014*** (0.001)	-0.013*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)
Physical health	-0.011*** (0.001)	-0.013*** (0.001)	-0.016*** (0.001)	-0.013*** (0.001)
Combined index	-0.012*** (0.001)	-0.013*** (0.001)	-0.009*** (0.001)	-0.008*** (0.001)
Grip strength	-0.019*** (0.002)	-0.010*** (0.001)	-0.025*** (0.002)	-0.013*** (0.001)

wght. stands for weighted

std. stands for standardized and for an gender-age indirect standardization

all cells show concentration indices

standard errors in parentheses

** p<0.05, *** p<0.01 indicating that indices differ significantly from zero

Table 4: Impact of SES on Subjective Health Measures Net of Objective Measure

				Generalized Ordered Probit					
	Worries: OLS	Low satisf.: OLS	SAH poor: OLS	SAH=1	SAH=2	SAH=3	SAH=4	SAH=5	Physical Health (SF12): OLS
Objective health measures									
SF12index	3.009***	2.625***	3.439***	-0.095***	-1.742***	-4.606***	6.063***	0.380***	0.1472***
(Column 9: grip strength)	(0.122)	(0.117)	(0.107)	(0.016)	(0.118)	(0.346)	(0.318)	(0.052)	(0.041)
SF12index*income	-0.269***	-0.257***	-0.113*	0.014	-0.097	-0.100	0.206	-0.023	-0.0001
(Column 9: grip strength)	(0.072)	(0.064)	(0.066)	(0.005)	(0.053)	(0.179)	(0.176)	(0.018)	(0.023)
Income									
Equhhinc	0.186*	0.169**	-0.010	0.003	-0.059	-0.147	0.178	0.025*	0.121**
	(0.103)	(0.084)	(0.099)	(0.008)	(0.080)	(0.221)	(0.215)	(0.014)	(0.054)
Equhhinc*equhhinc	-0.001	-0.001*	-0.000	-0.001**	-0.000	0.009***	-0.008***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.002)	(0.002)	(0.000)
Education									
9 years of schooling (d)	-0.058	-0.007	-0.011	-0.001	-0.033	-0.137*	0.171**	0.000	0.026
	(0.036)	(0.029)	(0.034)	(0.024)	(0.080)	(0.077)	(0.003)	(0.016)	
10 years of schooling (d)	-0.083**	0.008	-0.041	-0.001	-0.048**	-0.056	0.105	0.001	0.033*
	(0.035)	(0.028)	(0.033)	(0.001)	(0.021)	(0.079)	(0.078)	(0.003)	(0.017)
13 years of schooling (d)	-0.093**	0.028	-0.017	0.001	-0.039*	-0.128	0.166**	0.001	0.026*
	(0.037)	(0.032)	(0.035)	(0.002)	(0.023)	(0.085)	(0.085)	(0.003)	(0.015)
College degree (d)	-0.049	-0.053**	-0.019	-0.000	-0.023	0.031	-0.005	-0.003	0.018
	(0.030)	(0.026)	(0.029))	(0.002)	(0.017)	(0.055)	(0.054)	(0.002)	(0.012)
9yearsSchooling*equhhinc	0.033*	0.009	0.002	0.001	0.022	0.091*	-0.112**	-0.000	-0.011
	(0.020)	(0.018)	(0.020)	(0.001)	(0.018)	(0.051)	(0.049)	(0.002)	(0.009)
10yearsSchooling*equhhinc	0.041**	0.008	0.028	0.001	0.041**	0.039	-0.080*	-0.000	-0.009
	(0.018)	(0.015)	(0.018)	(0.001)	(0.018)	(0.050)	(0.048)	(0.002)	(0.009)
13yearsSchooling*equhhinc	0.038**	-0.004	0.017	-0.000	0.039**	0.056	-0.095*	-0.000	-0.003
	(0.017)	(0.015)	(0.017)	(0.001)	(0.018)	(0.051)	(0.050)	(0.001)	(0.008)
Other SES characteristics									
Full-time employed	-0.045**	-0.024	-0.047***	0.006*	-0.022*	-0.077**	0.094***	0.000	0.006
	(0.020)	(0.016)	(0.018)	(0.003)	(0.013)	(0.037)	(0.037)	(0.002)	(0.010)

Continued on next page...

	Worries: OLS	Low satisf.: OLS	SAH poor: OLS	Generalized Ordered Probit					Physical Health (SF12): OLS
				SAH=1	SAH=2	SAH=3	SAH=4	SAH=5	
Body Height in cm	-0.001 (0.001)	0.000 (0.001)	-0.001 (0.001)	-0.000 (0.000)	-0.001 (0.001)	-0.003 (0.002)	0.003 (0.002)	0.000 (0.000)	0.002*** (0.001)
Weight in kg	0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)	0.000 (0.000)	0.000 (0.000)	0.002 (0.002)	-0.002 (0.002)	-0.000 (0.000)	-0.002*** (0.000)
Full-time employed*equhhinc	0.016* (0.009)	0.007 (0.008)	0.019** (0.009)	-0.003*** (0.001)	0.009 (0.007)	0.018 (0.018)	-0.023 (0.018)	-0.001 (0.001)	-0.001 (0.005)
Married*equhhinc	0.006 (0.009)	0.005 (0.008)	0.021*** (0.008)	0.000 (0.001)	0.022*** (0.007)	-0.016 (0.018)	-0.008 (0.018)	0.002** (0.001)	-0.002 (0.005)
Kids*equhhinc	-0.024** (0.011)	0.000 (0.009)	-0.017 (0.011)	0.004*** (0.001)	-0.008 (0.012)	0.001 (0.026)	0.011 (0.024)	-0.001 (0.001)	0.005 (0.005)
R-squared	0.302	0.286	0.437			0.3587			0.3450
N	17678	17678	17678			17678			3855

* p<0.1, ** p<0.05, *** p<0.01

(d) for discrete change of dummy variable from 0 to 1

Marginal effects, which are calculated at the means of the covariates, are displayed

The following insignificant controls were included but are not displayed: immigrant, married, kids, East Germany, apprenticeship, apprenticeship*equhhinc, college*equhhinc, height*equhhinc, weight*equhhinc,immigrant*equhhinc, EastGermany*equhhinc

Equhhinc stands for equalized net household income divided by 1,000.

All regressions are weighted with sampling weights; original sample size is 19,529 (4,311 for column (10))

Reference categories are drop-out, non-full-time worker, mage1, fage1, fage1*equhhinc, mage1*equhhinc

In columns (1)-(3), the health measure, which is controlled for, is 1 minus sfl2index

In columns (1),(2), and (3), two out of eight gender-age dummies as well as two gender-age-income interactions are significant, respectively

In column (5), one gender-age dummy as well as one gender-age-income interaction are significantly different from zero

In columns (6),(7),(8), and (9) eight, eight, four, and seven gender-age dummies are significantly different from zero, respectively

In columns (6),(7),(8), and (9) three, three, three, and one gender-age-income interactions are significantly different from zero, respectively

Appendix A

The following health measures were collected in the 2006 SOEP questionnaire:

Subjective Health Measures

1. *What is your attitude towards the following areas – are you concerned about them?*
... Your health
 - (a) Very concerned
 - (b) Somewhat concerned
 - (c) Not concerned at all
2. How satisfied are you today with the following areas of your life?
Please answer by using the following scale: 0 means “totally unhappy,” 10 means “totally happy.”
How satisfied are you with your health?
0...10
3. *How would you describe your current health?*
 - (a) Very good
 - (b) Good
 - (c) Satisfactory
 - (d) Poor
 - (e) Bad

Use of health services

1. *Have you gone to a doctor within the last three months? If yes, please state how often.*
 - (a) Number of trips to the doctor’s in the last three months: X
 - (b) I haven’t gone to the doctor’s in the last three months.
2. *How many nights altogether did you spend in the hospital last year?*
 - (a) X nights
3. *If you were employed in 2005: How many days were you not able to work in 2005 because of illness? Please state all the days, not just those for which you had an official note from your doctor.*

- (a) None.
- (b) A total of X days.

Quasi Objective Health Measures and Calculation of SF12 Index

Andersen et al. (2007) provide a detailed explanation of how the physical and mental health components of the SF-12v2 were calculated and which German-specific algorithm was applied. A general overview of the questions included can be found online at <http://www.sf-36.org/demos/SF-12v2.html>.

The questions that were asked in the SOEP in 2006 differ slightly. The wordings and the original questionnaire can be found under “questionnaires” at <http://panel.gsoep.de/soepinfo2007/>.

Grip Strength

Hank et al. (2009) provide more details about the grip strength measurement in the SOEP survey. The SOEP Data Documentation series provides additional information including the original questionnaire with all grip strength-related questions and variables (Schupp, 2007). Figure 7 shows how the grip strength measurement was carried out.

Figure 7: Grip Strength Measurement



Appendix B

Table 5: Testing Differences of Concentration Indices and Dominance of Concentration Curves

Subjective (dichotomized)		Subjective (scaled)			Service Usage			(Quasi-)Objective				
Worries	Low Sat.	Poor SAH	SF12	HUI-III	15D	doctor	hospital	sickness	Mental (SF12)	Physical (SF12)	SF12-Index	Gripstrength
Subjective measures (dichotomized)												
Health worries		***	***	***	***	*	***		***	***	***	***
Low health satisfaction	c.d.		***	***	***	***		***	***	***	***	***
Poor SAH	c.d.	n.d.	***	***	***	***		***	***	***	***	***
Subjective measures (scaled)												
SAH (SF12-scaled)	c.d.	c.d.		***	***		***	***	***	***	***	***
HUI-III (SF12-scaled)	c.d.	c.d.	r.d.		***	***	**	*	***	***	***	***
15D (SF12-scaled)	c.d.	c.d.	r.d.	c.d.		***	***	*	***	***	***	***
Health System												
Doctor visits	c.d.	c.d.	c.d.	r.d.	c.d.		***	***				
Days in hospital	c.d.	n.d.	n.d.	r.d.	r.d.	n.d.		***	***	***	***	***
Sick leave days	c.d.	n.d.	n.d.	r.d.	r.d.	r.d.	n.d.		***	***	***	***
(Quasi-)Objective												
Mental health	c.d.	c.d.	c.d.	c.c.	c.d.	c.d.	c.d.	c.d.		*	***	***
Physical health	c.d.	c.d.	c.d.	c.d.	c.d.	c.d.	c.d.	c.d.	c.d.		**	**
SF12-Index	c.d.	c.d.	c.d.	c.c.	c.d.	c.d.	c.d.	c.d.	c.c.	c.c.		***
Gripstrength	c.d.	c.d.	c.d.	c.d.	c.d.	c.d.	n.d.	c.d.	n.d.	n.d.	c.d.	

* p<0.1, ** p<0.05, *** p<0.01 indicating differences in means between concentration indices

c.d. stands for “column dominates” meaning the concentration curve of the column’s health measure dominates the concentration curve of the row’s health measure

r.d. stands for “row dominates” meaning the concentration curve of the row’s health measure dominates the concentration curve of the columns’s health measure

n.d. stands for “non-dominance”

c.c. stands for “concentration curves cross”